

\*\*\*\*\*

\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 12 08:04:32 2000; MasPar time 13.26 Seconds  
Tabular output not generated. 566.101 Million cell updates/sec

Title: >US-09-276-268-20  
Description: (1-317) from US09276268.pep  
Perfect Score: 2378  
Sequence: 1 MRSGALWPLLWGALVTVGS.....KKAAGAGKVKSAQAKQAK 317

Scoring table: PAM 150  
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:geneseqp

Statistics: Mean 33.327; Variance 130.318; scale 0.256

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | ID  | Description             | Pred. No. |
|------------|-------|-------------|-----|-------------------------|-----------|
| 1          | 1575  | 66.2        | 263 | Human follistatin-3 pr  | 6.11e-155 |
| 2          | 796   | 33.5        | 319 | Follistatin.            | 1.22e-70  |
| 3          | 789   | 33.2        | 317 | Human follistatin (hFS) | 6.81e-70  |
| 4          | 785   | 33.0        | 304 | Rat EDF-binding protel  | 1.82e-69  |
| 5          | 783   | 32.9        | 304 | Human EDF-binding prot  | 2.99e-69  |
| 6          | 778   | 32.7        | 304 | Pig EDF-binding protel  | 1.02e-68  |
| 7          | 510   | 21.4        | 462 | Human polypeptide chai  | 2.54e-40  |
| 8          | 506   | 21.3        | 462 | Elongation factor-lalp  | 6.65e-40  |
| 9          | 496   | 20.9        | 398 | Prostate tumour induci  | 7.40e-39  |
| 10         | 428   | 18.0        | 460 | Plant elongation facto  | 8.79e-32  |
| 11         | 393   | 16.5        | 293 | Yeast elongation facto  | 3.59e-28  |
| 12         | 393   | 16.5        | 447 | Plant elongation facto  | 3.59e-28  |
| 13         | 392   | 16.5        | 458 | 61-1 clone polypeptide  | 3.59e-28  |
| 14         | 388   | 16.3        | 294 | Translation elongation  | 4.55e-28  |
| 15         | 388   | 16.3        | 448 | Plant elongation facto  | 1.17e-27  |
| 16         | 388   | 16.3        | 448 | LeEF-1 protein.         | 1.17e-27  |
| 17         | 388   | 16.3        | 448 | LeEF-1 clone A polypep  | 1.17e-27  |
| 18         | 188   | 7.9         | 664 | Plant elongation facto  | 1.17e-27  |
| 19         | 170   | 7.1         | 308 | Human HP-8 antigen ass  | 8.98e-08  |
| 20         | 168   | 7.1         | 380 | Follistatin related pr  | 4.28e-06  |
| 21         | 160   | 6.7         | 499 | Transforming growth fa  | 6.55e-06  |
| 22         | 143   | 6.0         | 303 | Human GSP1-rk.          | 3.55e-05  |
| 23         | 140   | 5.9         | 374 | Sequence of human bone  | 1.21e-03  |
|            |       |             |     | Human transforming gro  | 2.23e-03  |

|    |     |     |      |   |        |                        |          |
|----|-----|-----|------|---|--------|------------------------|----------|
| 24 | 140 | 5.9 | 374  | 1 | W37497 | Human TMP-2.           | 2.23e-03 |
| 25 | 133 | 5.6 | 500  | 1 | W94496 | Human delta-2 protein  | 9.20e-03 |
| 26 | 133 | 5.6 | 659  | 1 | W94497 | Human delta-2 mature p | 9.20e-03 |
| 27 | 133 | 5.6 | 685  | 1 | W80813 | Nucleotide sequence of | 9.20e-03 |
| 28 | 133 | 5.6 | 685  | 1 | W80813 | Human delta-2 protein. | 9.20e-03 |
| 29 | 129 | 5.4 | 44   | 1 | W56633 | HP-8 auto-antigen, rel | 2.05e-02 |
| 30 | 129 | 5.4 | 44   | 1 | W04328 | Human HP-8 antigen epi | 2.05e-02 |
| 31 | 125 | 5.3 | 51   | 1 | W04330 | Human HP-8 antigen epi | 4.54e-02 |
| 32 | 127 | 5.3 | 86   | 1 | W62074 | Human tumour-associate | 3.05e-02 |
| 33 | 127 | 5.3 | 86   | 1 | W15067 | Human PEC-60-type prot | 3.05e-02 |
| 34 | 122 | 5.1 | 104  | 1 | R36608 | Thrombin inhibitory pr | 8.19e-02 |
| 35 | 118 | 5.0 | 282  | 1 | R79101 | Prostaglandin I2 (PGI2 | 1.79e-01 |
| 36 | 118 | 5.0 | 282  | 1 | R79102 | Prostaglandin I2 (PGI2 | 1.79e-01 |
| 37 | 118 | 5.0 | 282  | 1 | Y06833 | Angiomodulin/tumour-de | 1.79e-01 |
| 38 | 118 | 5.0 | 548  | 1 | R33741 | XRI.                   | 1.79e-01 |
| 39 | 106 | 4.5 | 436  | 1 | W74722 | Human secreted protein | 1.78e+00 |
| 40 | 106 | 4.5 | 873  | 1 | R44735 | apo-E lipoprotein rece | 1.78e+00 |
| 41 | 108 | 4.5 | 1801 | 1 | W50895 | Rat laminin B2 chain.  | 1.22e+00 |
| 42 | 108 | 4.5 | 1964 | 1 | W95557 | Mus musculus notch4 pr | 1.22e+00 |
| 43 | 105 | 4.4 | 612  | 1 | W39256 | Human partial mature m | 2.15e+00 |
| 44 | 105 | 4.4 | 846  | 1 | R74692 | Rat very low density 1 | 2.15e+00 |
| 45 | 105 | 4.4 | 1251 | 1 | R79475 | Mouse LTBP-3.          | 2.15e+00 |

## ALIGNMENTS

RESULT 1  
ID Y01098 standard; Protein; 263 AA.  
AC Y01098;  
DT 11-JUN-1999 (first entry)  
DE Human follistatin-3 protein sequence.  
KW Follistatin-3; human; cancer; cellular growth disorder; liver cirrhosis;  
KW differentiation disorder; reproductive system disorder; male sterility;  
KW activin-induced differentiation inhibitor; gonadotroph adenoma; hepatoma;  
KW osteosarcoma; idiopathic pulmonary fibrosis; pulmonary fibrosis; tumour;  
KW fibrotic disorder; osteoarthritis; haematopoiesis; infectious disease;  
KW sepsis; cancer; sillcosis; sarcoidosis; endotoxemic shock; therapy.  
OS Homo sapiens.  
PN W09910364-Al.  
PD 04-MAR-1999.  
PF 27-AUG-1998: U17710.  
PR 29-AUG-1997: US-056248.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Duan R, Ruben SM;  
DR WPI; 99-204646/17;  
DR N-PSDB; X28124.  
PT New follistatin-3 polypeptides and nucleic acids - used to develop  
PT products for treating e.g. cancers, male sterility, wound healing,  
PT fibrotic disorders, angiogenesis and autoimmune, inflammatory and  
PT infective diseases  
PS Claim 18; Fig 1; 109pp; English.  
CC This sequence is the follistatin-3 (FS3) protein of the invention.  
CC The products can be used to treat cancers and other cellular growth and  
CC differentiation disorders as well as disorders of the reproductive  
CC system. FS3 can be used or to treat male sterility. FS3 may also be used  
CC to inhibit the activin-induced differentiation of follicular granulosa  
CC cells. FS3 may be used therapeutically to regulate autocrine endothelial  
CC cell activity and, as a result, induce angiogenesis. Treatment to  
CC increase the expression of the presence of FS3 may be used to inhibit the  
CC progression of gonadotroph adenomas, osteosarcomas, hepatomas, and other  
CC tumours and cancers. FS3 may also be used to treat other fibrotic  
CC disorders including liver cirrhosis, osteoarthritis and pulmonary  
CC fibrosis. It may also be used to regulate haematopoiesis, and to treat  
CC sepsis. Antagonists of FS3 may be used to treat a deficiency in FSH,  
CC oestrogen and other hormones, to prevent or inhibit or reduce the  
CC production of spermatozoa, to modulate gonadal androgen biosynthesis. FS3  
CC antagonists may also be used to treat infectious diseases including  
CC sillcosis, sarcoidosis, idiopathic pulmonary fibrosis by altering the  
CC activation state of mononuclear phagocytes, to treat idiopathic  
CC hyper-eosinophilic syndrome by preventing eosinophil production and  
CC activation. Endotoxemic shock may also be treated by FS3 antagonists by  
CC preventing the activation of macrophages.  
SQ Sequence 263 AA;

Query Match 66.2%; Score 1575; DB 1; Length 263;  
 Best Local Similarity 79.4%; Pred. No. 6.11e-155;  
 Matches 201; Conservative 27; Mismatches 23; Indels 2; Gaps 2;

Db 6 PGPLWPLPWGALAWAVGVSS-MGSGNPAGVGVWLGQGGATCSLVLTQDVTRAECAS 64  
 QY 3 SGALWPLWLGALVWTVGVAVGMSVSGVGVWLGQGGATCSLVLTQDVTRAECAS 62

Db 65 GNIDTANSNLTHPGKNLILGLFLGLVCLPCKDCSDGVEGPGKACRMGLGGRPR-CECAP 123  
 QY 63 GNIDTANSNLTHPGKNLILGLFLGLVCLPCKDCSDGVEGPGKACRMGLGGRPR-CECAP 122

Db 124 DCSGLPARLVQCSGDATYDECELRARCRGHPDLVYMYRGCRKSCHEVWVCPRPQSCV 183  
 QY 123 NCEGFPAGVQVCGSDATYDECELRARCRGHPDLVYMYRGCRKSCHEVWVCPRPQSCV 182

Db 184 VDQTGSAHCVVCRAPCPVPSPPQELCGNNNTYISSCHMRQATCFLGRSIVGRHAGSC 243  
 QY 183 VDQTGSAHCVVCRAPCPVPSPPQELCGNNNTYISSCHMRQATCFLGRSIVGRHAGSC 242

Db 244 AGTPEPPGESA 256  
 QY 243 TGGPKFLKSGDAA 255

RESULT 2  
 ID R74600 standard; protein; 319 AA.  
 AC R74600;  
 DE 24-OCT-1995 (first entry)  
 KW Neuronal differentiation; transforming growth factor; TGF-beta;  
 KW degenerative disorders; nervous system; Alzheimer's disease;  
 KW Parkinson's disease; amyotrophic lateral disease; Pick's disease;  
 KW Huntington's disease; multiple sclerosis; anoxia-ischaemia; trauma;  
 KW ageing; tachycardia; atrial cardiac arrhythmia.  
 OS Synthetic.  
 PN W09510611-A.  
 PD 20-APR-1995.  
 PF 14-OCT-1995; 117450.  
 PR 14-OCT-1993; US-136748.  
 PA (HARD) Harvard College.  
 PI Melton D, Hemmati-Brivanlou A;  
 DR WPI; 95-178527/23.  
 DR N-PSDB; Q90017.

Inducing neuronal differentiation using TGF-beta family agonist  
 useful to prevent death and/or degeneration of neuronal cells, e.g.  
 to treat Alzheimer's disease, multiple sclerosis, etc.  
 PS Example; Page 41; 53pp; English.  
 CC The sequence is that of follistatin.  
 CC differentiate neuronal cell phenotype by antagonising the action of  
 CC at least one member of the transforming growth factor beta (TGF-beta)  
 CC family which normally induces the cell to differentiate to a non-  
 CC neuronal phenotype. The agent can be used to prevent death and/or  
 CC degeneration of a neuronal cell in a degenerative disorder of the  
 CC nervous system, e.g. Alzheimer's disease, Parkinson's disease,  
 CC amyotrophic lateral sclerosis, Pick's disease, Huntington's disease,  
 CC multiple sclerosis, neuronal damage resulting from anoxia-ischaemia  
 CC or trauma and neuronal degeneration associated with ageing.  
 SQ Sequence 319 AA;

Query Match 33.5%; Score 796; DB 1; Length 319;  
 Best Local Similarity 46.0%; Pred. No. 1.22e-70;  
 Matches 99; Conservative 46; Mismatches 57; Indels 3; Gaps 3;

Db 26 AVQAGNWLQSNKRGQVLYRTLSKECKTGRLTSTWTEEDVPNSTLTKWIFHGGA 85  
 QY 30 SVPGGVWLGQGGATCSLVLTQDVTRAECASGNTANSNTHPGNKI-SLGLGFLV 88

Db 86 PHCIPCKETCENVDCGPGKCKMKNKPRCVCAPDCSNITWKGSCGIDGTYKDECAL 145  
 QY 89 -HCLPKDCSDGVEGPGKACRMAGNAGSNCEVPCGEPAGVQVCGSDGATYRDECEL 147

Db 146 LKAKCKGVPDLVQYQKCKKTKRDVLCPGSSCVVDQTNNAVCVTCNRI-CPEPTSPDQ 204  
 QY 148 RTARCRGHGPDLRVYRGCRKSCAQVCPRPQSCVLDQVTSAGHCVVCRAPCPVPSNPQ 207

Db 205 YLCGNDGITYGSACHLRKATCLLGRSISGLAYEGKC 239  
 QY 208 ELCGNNNTYISSCHLRQATCFLGRSIVGRHAGSC 242

RESULT 3  
 ID P93396 standard; protein; 317 AA.  
 AC P93396;  
 DE 15-JUL-1990 (first entry)  
 KW Human follistatin (hFS) precursor protein as encoded by cDNA  
 KW Human follistatin precursor protein; signal sequence;  
 KW testicular lambda gtl1 cDNA library; porcine follistatin precursor.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 4..5 /note="In pFS this is PK"  
 FT misc\_difference 163 /note="in pFS this is K"  
 FT misc\_difference 200 /note="in pFS this is T"  
 FT misc\_difference 276 /note="in pFS this is E"  
 FT misc\_difference 280 /note="in pFS this is E"  
 FT region 124 /note="in pFS this is E"  
 FT region 258 /note="potential N-linked glycosylation site"  
 FT peptide 1..29 /note="potential N-linked glycosylation site"  
 FT /note="signal peptide" 30..317

protein W08901945-A.  
 PN 09-MAR-1989.  
 PD 26-AUG-1988; U02971.  
 PR 26-JUL-1988; US-224432.  
 PA (SALK) Salk Inst for Biol Stud.  
 PI Ling NCK, Ueno N, Shimasaki S, Ying SY, Guillemin RCL;  
 DR WPI; 89-085524/11.  
 DR N-PSDB; N90641.  
 PT Monomeric mammalian follistatin protein -  
 PT isolated from porcine follicular fluid and prepd. by recombinant  
 PT techniques, used esp. as male contraceptive  
 PS Disclosure; 34pp; English.  
 CC A testicular lambda gtl1 cDNA library was screened with a cDNA probe  
 CC encoding the first 317 amino acids of the porcine FS (pFS) precursor.  
 CC 12 positive clones were obtained. From the nucleotide sequences of  
 CC the clones, it was determined that there is encoded an hFS precursor  
 CC containing a 344 amino acid sequence which is highly homologous with  
 CC the 344-residue pFS precursor. It differs from the 344-residue pFS  
 CC precursor by only 6 residues (see FT). p93396 is encoded by bases 28-978  
 CC of n90641.  
 SQ Sequence 317 AA;

Query Match 33.2%; Score 789; DB 1; Length 317;  
 Best Local Similarity 45.6%; Pred. No. 6.81e-70;  
 Matches 99; Conservative 46; Mismatches 69; Indels 3; Gaps 3;

Db 24 DRSAQAGNWLQSNKRGQVLYRTLSKECKTGRLTSTWTEEDVPNSTLTKWIFHGGA 83  
 QY 28 EDSPGVWLGQGGATCSLVLTQDVTRAECASGNTANSNTHPGNKI-SLGLGFLV 86

Db 84 GAPNCPCKETCENVDCGPGKCKMKNKPRCVCAPDCSNITWKGSCGIDGTYRNEC 143  
 QY 87 LV-HCLPKDCSDGVEGPGKACRMAGNAGSNCEVPCGEPAGVQVCGSDGATYRDEC 145

Db 144 ALLKARCKEPELEVOYQKCKKTKRDVLCPGSSCVVDQTNNAVCVTCNRI-CPEPASS 202  
 QY 146 ELRTARCRGHGPDLRVYRGCRKSCAQVCPRPQSCVLDQVTSAGHCVVCRAPCPVPSNP 205

\*\*\*\*\*

WPELEH (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Wed Apr 12 08:05:05 2000; MasPar time 8.58 Seconds  
Tabular output not generated. 478.795 Million cell updates/sec

Title: >US-09-276-268-20  
Description: (1-317) from US09276268.pep  
Perfect Score: 2378  
Sequence: 1 MRSGALWPLMGALVWTVGSG.....KKAAGKVKTSQAQKAK 317

Scoring table: PAM 150  
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1  
Statistics: Mean 31.424; Variance 133.180; scale 0.236

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description                       | Pred. No. |
|------------|-------|---------------|--------|----|-----------------------------------|-----------|
| 1          | 1575  | 66.2          | 263    | 2  | US-08-972-Sequence 2, Applicatio  | 1.03e-143 |
| 2          | 519   | 21.8          | 73     | 2  | US-08-972-Sequence 5, Applicatio  | 1.41e-38  |
| 3          | 510   | 21.4          | 462    | 4  | 5253348-1 Patent No. 5225348.     | 1.06e-37  |
| 4          | 510   | 21.4          | 462    | 2  | US-08-371-Sequence 18, Applicati  | 1.06e-37  |
| 5          | 506   | 21.3          | 462    | 1  | US-08-299-Sequence 1, Applicatio  | 2.58e-37  |
| 6          | 488   | 20.5          | 397    | 2  | US-08-371-Sequence 19, Applicati  | 1.43e-35  |
| 7          | 483   | 20.3          | 398    | 2  | US-08-371-Sequence 17, Applicati  | 4.37e-35  |
| 8          | 453   | 19.0          | 71     | 2  | US-08-972-Sequence 4, Applicatio  | 3.45e-32  |
| 9          | 261   | 11.0          | 1940   | 2  | US-08-644-Sequence 30, Applicati  | 4.97e-14  |
| 10         | 188   | 7.9           | 664    | 1  | US-08-421-Sequence 6, Applicatio  | 1.91e-07  |
| 11         | 160   | 6.7           | 499    | 2  | US-08-820-Sequence 40, Applicati  | 5.18e-05  |
| 12         | 143   | 6.0           | 303    | 4  | 5340934-13 Patent No. 5340934.    | 1.42e-03  |
| 13         | 140   | 5.9           | 374    | 2  | US-08-820-Sequence 25, Applicatio | 2.53e-03  |
| 14         | 129   | 5.4           | 44     | 1  | US-08-421-Sequence 4, Applicatio  | 2.05e-02  |
| 15         | 125   | 5.3           | 51     | 1  | US-08-421-Sequence 2, Applicatio  | 4.33e-02  |
| 16         | 127   | 5.3           | 86     | 2  | US-08-744-Sequence 1, Applicatio  | 2.98e-02  |
| 17         | 127   | 5.3           | 86     | 2  | US-08-149-Sequence 1, Applicatio  | 2.98e-02  |
| 18         | 122   | 5.1           | 103    | 1  | US-08-211-Sequence 3, Applicatio  | 7.57e-02  |
| 19         | 122   | 5.1           | 184    | 1  | US-08-211-Sequence 7, Applicatio  | 7.57e-02  |
| 20         | 122   | 5.1           | 184    | 1  | US-08-211-Sequence 15, Applicati  | 7.57e-02  |
| 21         | 122   | 5.1           | 186    | 1  | US-08-211-Sequence 9, Applicatio  | 7.57e-02  |
| 22         | 122   | 5.1           | 368    | 1  | US-08-211-Sequence 17, Applicati  | 7.57e-02  |
| 23         | 122   | 5.1           |        |    |                                   |           |

|    |     |     |      |   |                                   |          |
|----|-----|-----|------|---|-----------------------------------|----------|
| 24 | 118 | 5.0 | 548  | 1 | US-08-463-Sequence 2, Applicatio  | 1.58e-01 |
| 25 | 118 | 5.0 | 548  | 1 | US-08-694-Sequence 2, Applicatio  | 1.58e-01 |
| 26 | 118 | 5.0 | 548  | 1 | US-08-333-Sequence 2, Applicatio  | 1.58e-01 |
| 27 | 116 | 4.9 | 129  | 1 | US-08-211-Sequence 11, Applicati  | 2.29e-01 |
| 28 | 115 | 4.8 | 108  | 1 | US-08-211-Sequence 13, Applicati  | 2.75e-01 |
| 29 | 105 | 4.4 | 1193 | 2 | US-08-400-Sequence 10, Applicati  | 1.67e+00 |
| 30 | 105 | 4.4 | 1231 | 3 | PCT-US95-0 Sequence 3, Applicatio | 1.67e+00 |
| 31 | 105 | 4.4 | 1252 | 2 | US-08-199-Sequence 3, Applicatio  | 1.67e+00 |
| 32 | 105 | 4.4 | 1252 | 2 | US-08-316-Sequence 3, Applicatio  | 1.67e+00 |
| 33 | 103 | 4.3 | 81   | 2 | US-09-149-Sequence 4, Applicatio  | 2.37e+00 |
| 34 | 103 | 4.3 | 81   | 2 | US-08-744-Sequence 4, Applicatio  | 2.37e+00 |
| 35 | 102 | 4.3 | 1015 | 1 | US-08-537-Sequence 1, Applicatio  | 2.83e+00 |
| 36 | 102 | 4.3 | 2471 | 1 | US-08-185-Sequence 16, Applicati  | 2.83e+00 |
| 37 | 102 | 4.3 | 2471 | 1 | US-08-083-Sequence 19, Applicati  | 2.83e+00 |
| 38 | 100 | 4.2 | 86   | 2 | US-09-149-Sequence 3, Applicatio  | 4.02e+00 |
| 39 | 100 | 4.2 | 86   | 2 | US-08-744-Sequence 3, Applicatio  | 4.02e+00 |
| 40 | 100 | 4.2 | 219  | 1 | US-08-152-Sequence 31, Applicati  | 4.02e+00 |
| 41 | 100 | 4.2 | 219  | 2 | US-08-125-Sequence 18, Applicati  | 4.02e+00 |
| 42 | 100 | 4.2 | 219  | 2 | US-08-460-Sequence 18, Applicati  | 4.02e+00 |
| 43 | 99  | 4.2 | 846  | 1 | US-08-451-Sequence 3, Applicatio  | 4.79e+00 |
| 44 | 99  | 4.2 | 873  | 1 | US-08-393-Sequence 2, Applicatio  | 4.79e+00 |
| 45 | 99  | 4.2 | 1833 | 3 | PCT-US95-0 Sequence 18, Applicati | 4.79e+00 |

ALIGNMENTS

|        |                 |
|--------|-----------------|
| RESULT | 1               |
| ID     | US-08-972-008-2 |
| XX     | STANDARD;       |
| AC     | xxxxxx          |
| DT     |                 |
| XX     |                 |
| DE     |                 |
| XX     |                 |

Sequence 2, Application US/08972008

Sequence 2, Application US/08972008  
Patent No. 5942420  
GENERAL INFORMATION:

APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: No. 5942420el Molecules of the Follistatin-Related  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/972,008  
FILING DATE:

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: MMI-026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 263 amino acids  
TYPE: amino acid

MOLECULE TYPE: protein  
SEQUENCE 263 AA: 27663 MW: 326569 CN;  
Query Match

Score 1575; DB 2; Length 263;  
66.2%;



\*\*\*\*\*  
WATERMAN  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Apr 12 07:46:37 2000; MasPar time 36.61 Seconds  
640.186 Million cell updates/sec

Tabular output not generated.

Title: >US-09-276-268-16  
Description: (1-338) from US09276268.pgp  
Perfect Score: 2401  
Sequence: 1 MGAVWSALLVGGGLAGLIL.....NRFLQTGHEDKVVHGWGIIH 338

Scoring table: PAM 150  
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmb12  
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human  
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle  
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified  
13:sp-vertebrate 14:sp-virus

Statistics: Mean 47.028; Variance 99.789; scale 0.471

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES  |       |             |                                  |           |
|------------|-------|-------------|----------------------------------|-----------|
| Result No. | Score | Query Match | Description                      | Pred. No. |
| 1          | 1381  | 57.5        | GENETHONIN 1.                    | 2.14e-238 |
| 2          | 163   | 6.8         | GLUCAMYLASE G2 (EC 3.1.28e-08    |           |
| 3          | 135   | 5.6         | ALPHA-AMYLASE (EC 3.2.1.37e-04   |           |
| 4          | 134   | 5.6         | COLLAGEN PRO ALPHA-CHA 1.89e-04  |           |
| 5          | 128   | 5.3         | MINI-COLLAGEN PRECURSOR 1.24e-03 |           |
| 6          | 128   | 5.3         | F9D12.19 PROTEIN. 1.24e-03       |           |
| 7          | 124   | 5.2         | F57B7.3 PROTEIN. 4.23e-03        |           |
| 8          | 125   | 5.2         | GLUCAMYLASE. 3.12e-03            |           |
| 9          | 123   | 5.1         | BETA-AMYLASE (EC 3.2.1.574e-03   |           |
| 10         | 123   | 5.1         | PRECOLLAGEN P PRECURSOR 5.74e-03 |           |
| 11         | 122   | 5.1         | TYPE I PROCOLLAGEN PRO 7.77e-03  |           |
| 12         | 122   | 5.1         | COLLAGEN ALPHA1 (FRAGM 7.77e-03  |           |
| 13         | 122   | 5.1         | TYPE I COLLAGEN PRO-AL 7.77e-03  |           |
| 14         | 122   | 5.1         | PRO ALPHA 1(I) COLLAG 7.77e-03   |           |
| 15         | 120   | 5.0         | TYPE II COLLOGEN CYANO 1.42e-02  |           |
| 16         | 119   | 5.0         | F58F6.2 PROTEIN. 1.91e-02        |           |
| 17         | 119   | 5.0         | TYPE II COLLOGEN CYANO 1.91e-02  |           |
| 18         | 119   | 5.0         | FIBRILLAR COLLAGEN (FR 1.91e-02  |           |
| 19         | 119   | 5.0         | CYCLODEXTRIN GLYCOSYLT 1.91e-02  |           |
| 20         | 120   | 5.0         | PGRS-FAMILY PROTEIN. 1.42e-02    |           |

|    |     |     |      |    |        |                        |          |
|----|-----|-----|------|----|--------|------------------------|----------|
| 21 | 120 | 5.0 | 890  | 5  | 077087 | FIBRILLAR COLLAGEN CHA | 1.42e-02 |
| 22 | 121 | 5.0 | 1160 | 4  | Q14046 | COL2A1 ALPHA1 (II) COL | 1.05e-02 |
| 23 | 120 | 5.0 | 1364 | 6  | Q62649 | ALPHA2(I) COLLAGEN.    | 1.42e-02 |
| 24 | 120 | 5.0 | 1418 | 13 | Q9W7R9 | ALPHA1 TYPE II COLLAG  | 1.42e-02 |
| 25 | 120 | 5.0 | 1418 | 6  | Q28396 | TYPE II COLLAGEN.      | 1.42e-02 |
| 26 | 120 | 5.0 | 1445 | 13 | Q93251 | ALPHA 1 TYPE I COLLAG  | 1.42e-02 |
| 27 | 121 | 5.0 | 1487 | 4  | Q14047 | ALPHA-1 TYPE II COLLAG | 1.05e-02 |
| 28 | 119 | 5.0 | 1497 | 11 | Q61431 | PROCOLLAGEN, TYPE V, A | 1.91e-02 |
| 29 | 121 | 5.0 | 1549 | 11 | Q60444 | TYPE VII COLLAGEN (FRA | 1.05e-02 |
| 30 | 120 | 5.0 | 2391 | 10 | Q9XE40 | HYPOTHETICAL 27.9 KD P | 1.42e-02 |
| 31 | 120 | 5.0 | 2944 | 11 | Q63870 | TYPE VII COLLAGEN.     | 1.42e-02 |
| 32 | 121 | 5.0 | 3942 | 11 | Q88737 | BASSON.                | 1.05e-02 |
| 33 | 117 | 4.9 | 639  | 3  | Q12537 | GLUCAMYLASE PRECURSOR  | 3.46e-02 |
| 34 | 118 | 4.9 | 940  | 4  | Q00405 | FB19 PROTEIN.          | 2.57e-02 |
| 35 | 117 | 4.9 | 1718 | 5  | Q62603 | HYPOTHETICAL 186.0 KD  | 3.46e-02 |
| 36 | 116 | 4.8 | 564  | 11 | Q70604 | COLLAGEN ALPHA 1 (III) | 4.64e-02 |
| 37 | 115 | 4.8 | 1650 | 11 | Q921W0 | COLLAGEN ALPHA-2 (XI). | 6.23e-02 |
| 38 | 114 | 4.7 | 287  | 5  | Q18480 | SIMILAR TO CUTICULAR C | 8.33e-02 |
| 39 | 114 | 4.7 | 289  | 5  | Q18799 | C53B4.5 PROTEIN.       | 8.33e-02 |
| 40 | 114 | 4.7 | 303  | 5  | Q23364 | SIMILAR TO CUTICULAR C | 8.33e-02 |
| 41 | 114 | 4.7 | 635  | 4  | Q15336 | ELASTIN (FRAGMENT).    | 8.33e-02 |
| 42 | 114 | 4.7 | 1419 | 11 | Q63123 | COLLAGEN ALPHA 1 TYPE  | 8.33e-02 |
| 43 | 113 | 4.7 | 1450 | 13 | Q9Y1B4 | ALPHA 1 TYPE I COLLAG  | 1.11e-01 |
| 44 | 113 | 4.7 | 1491 | 13 | Q91718 | ALPHA-1 TYPE II' COLLA | 1.11e-01 |
| 45 | 114 | 4.7 | 1655 | 5  | Q24754 | MASTERMIND.            | 8.33e-02 |

ALIGNMENTS

| RESULT                | ID  | PRELIMINARY; | PRT; | 358 AA. |
|-----------------------|---|--------------|------|---------|
| AC                    | O95210;   |              |      |         |
| DT                    | 01-MAY-1999 (TREMblrel. 10, Created)                                  |              |      |         |
| DT                    | 01-MAY-1999 (TREMblrel. 10, Last sequence update)                     |              |      |         |
| DT                    | 01-MAY-1999 (TREMblrel. 10, Last annotation update)                   |              |      |         |
| DE                    | GENETHONIN 1.   |              |      |         |
| OS                    | Homo sapiens (Human).   |              |      |         |
| OC                    | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;         |              |      |         |
| OC                    | Eutheria; Primates; Catarrhini; Homnidae; Homo.                       |              |      |         |
| RN                    | [1]   |              |      |         |
| RP                    | SEQUENCE FROM N.A.  |              |      |         |
| RC                    | TISSUE=MUSCLE;  |              |      |         |
| RX                    | MEDLINE; 96425696.  |              |      |         |
| RA                    | PIETU G., ALIBERT O., GUICHARD V., LAWY B., BOIS F., LEROY E.,        |              |      |         |
| RA                    | MARIAGE-SAMPSON R., HOULGATTE R., SOULARIE P., AUFRAY C.;             |              |      |         |
| RT                    | "Novel gene transcripts preferentially expressed in human muscles     |              |      |         |
| RT                    | revealed by quantitative hybridization of a high density cDNA array." |              |      |         |
| RL                    | Genome Res. 6:492-503(1996).  |              |      |         |
| RN                    | [2]   |              |      |         |
| RP                    | SEQUENCE FROM N.A.  |              |      |         |
| RC                    | TISSUE=MUSCLE;  |              |      |         |
| RX                    | MEDLINE; 99013693.  |              |      |         |
| RA                    | BOUJY S., LIGNON M.-F., PIETU G., LE CUNFF M., LEGER J.-J.,           |              |      |         |
| RA                    | AUFRAY C., DECHESENE C.A.;  |              |      |         |
| RT                    | "Molecular cloning and functional expression of a novel human gene    |              |      |         |
| RT                    | encoding two 41-43 kDa skeletal muscle internal membrane proteins."   |              |      |         |
| RL                    | Biochem. J. 335:549-556(1998).  |              |      |         |
| DR                    | EMBL; AF062534; AAC78827.1; -.  |              |      |         |
| SQ                    | SEQUENCE 358 AA; BBC512D6 CRC32;                                      |              |      |         |
| Query Match           | 57.5%; Score 1381; DB 4; Length 358;                                  |              |      |         |
| Best Local Similarity | 62.6%; Pred. No. 2.14e-238;   |              |      |         |
| Matches               | 226; Conservative 53; Mismatches 56; Indels 26; Gaps 16;              |              |      |         |
| Db                    | 1 MGAVWSALLVGGGLAGLILFVWLLRGGPGDTCGDGDAQ-EKADPLGGAALPG-CHOSGSS 58     |              |      |         |
| QY                    | 1 MGAVWSALLVGGGLAGLILWLLRGGDSGAGKGVAPPPGKAPGGAAPGDPGGGGS 60           |              |      |         |
| Db                    | 59 G-LSPPGSGGELVTKPEHLQESNGHLISKTKDLGKLAQASWRQNPSREYCDNSRHPVS 117     |              |      |         |
| QY                    | 61 GGLSPESDRELVSKEHLRESNGHLISESKDGLNLPDAQ-RLQNVGADWV-NAREFVPV 118     |              |      |         |
| Db                    | 118 GQFPDTEPATSETNSRSYSSEVSRSNLSLESPMGENGWFGQGEISAKAATCAEKLPPSN 177   |              |      |         |

QY 119 KIPDTHSRADSEAA--RN--Q-SPG-S-HG--GEWRLPKGQETAVKVGSAVAAKLASSS 169  
 Db 178 LLKNRAEEMSLDLSQDRVDHEWEMVPRHSHGVDGVGSLKAPVLNMQGMDNGRS 237  
 QY 170 LLVDRAR-----A-V-SQDQAGHEDWEVSRHSHGVSGLGSLASRLSLNQRMDDSTN 222  
 Db 238 TLVEARGQGVHGMKERNVAVPAGSQOVSVRFOVHYVSTVDQVFAVTCGHECLGRWNTYI 297  
 QY 223 SLVGRGWEVDGKVA--SLKP---QOVSIOFQVHYTNTDQVFAVTCGHECLGRWNTYI 277  
 Db 298 PLHYNKDGFWSHSLFLPADTVVWKNFLVGVNGVTRWECSNRFLQTHGHEKVVHAWNGI 357  
 QY 278 PLHYCKDGLWSHSLFLPADTVVWKNFLVGVNGVTRWECSNRFLQTHGHEKVVHAWNGI 337  
 Db 358 H 358  
 QY 338 H 338

RESULT 2  
 ID Q12596; PRELIMINARY; PRT; 579 AA.  
 AC Q12596;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE GLUCOAMYLASE G2 (EC 3.2.1.3).  
 OS Corticium rolfsii.  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Stereales;  
 OC Corticiaceae; Corticium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AHU9627;  
 RX MEDLINE; 96158471.  
 RA NAGASAKA Y., MURAKI N., KIMURA A., SUTO M., YOKOTA A., TOMITA F.;  
 RT "Cloning of Corticium rolfsii glucoamylase cDNA and its expression in  
 RT Saccharomyces cerevisiae.";  
 RL Appl. Microbiol. Biotechnol. 44:451-458(1995).  
 DR EMBL; D49448; BAA08436.1; .  
 DR HSSP; P04064; IGLM.  
 DR PFAM; PF00686; CBD.4; 1.  
 DR PFAM; PF00723; Glyco\_hydro.15; 1.  
 DR PRINTS; PR00736; GLHYDLASE15.  
 KW Hydrolase; Glycosidase.  
 SQ SEQUENCE 579 AA; 61548 MW; 0439FA95 CRC32;

Query Match 6.8%; Score 163; DB 3; Length 579;  
 Best Local Similarity 36.3%; Pred. No. 1.28e-08;  
 Matches 29; Conservative 16; Mismatches 30; Indels 5; Gaps 5;  
 Db 483 VEVTFDV-YATTYVGNIIYITGDVSELGNTPANGVALSSANTYPTWSATIALPADTTIQY 541  
 QY 244 VSTQFQVHYTNTDQVFIATVGDHESLGRN-TY-IPLHYCKDGLWSHSLFLPADTVVW 301  
 Db 542 KYVNDGSTVI-WEDAIISR 560  
 QY 302 KEVLVENKEVTRWEEC-SNR 320

RESULT 3  
 ID Q59222; PRELIMINARY; PRT; 613 AA.  
 AC Q59222;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
 DE ALPHA-AMYLASE (EC 3.2.2.1) (PURINE NUCLEOSIDASE).  
 GN AMY.  
 OS Bacillus sp.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TS-23;  
 RX MEDLINE; 92064646.

RA LIN L.L., CHU W.S., HSU W.H.;  
 CC Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: A N-D-RIBOSYLPURINE + H(2)O -> A PURINE + D-  
 CC RIBOSE.  
 DR EMBL; U22045; AAA63900.1; .  
 DR HSSP; P06278; 1BPL.  
 DR PFAM; PF00128; alpha-amylose; 1.  
 DR PFAM; PF00686; CBD\_4; 1.  
 KW Hydrolase; Glycosidase.  
 SQ SEQUENCE 613 AA; 69536 MW; DC66587C CRC32;

Query Match 5.6%; Score 135; DB 2; Length 613;  
 Best Local Similarity 27.7%; Pred. No. 1.37e-04;  
 Matches 31; Conservative 17; Mismatches 60; Indels 4; Gaps 4;  
 Db 483 LTGNRSDVTNTNADGGEFKVNGSVSIWAKTSNVTFTVNNATTSGQNVVYVANIPEL 542  
 QY 213 LNQRMDSTNSLVGGRG-WEVDGKVASLKPOQVS-IQFQVHYTNTDQVFIATVGDHESL 270  
 Db 543 GNWNTANAIAKMPSSYPTWKATIALPOGKAIEFKKDKQAGNVIWESTSNR 594  
 QY 271 GRWNT-Y-IPLHYCKDGLWSHSLFLPADTVVWKNFLVGVNGVTRWECSNR 320

RESULT 4  
 ID O97405; PRELIMINARY; PRT; 1378 AA.  
 AC O97405;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
 DE COLLAGEN PRO ALPHA-CHAIN PRECURSOR.  
 GN HDCOL 1 ALPHA.  
 OS Eukaryota; Metazoa; Mollusca; Gastropoda; Archaeogastropoda;  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Archaeogastropoda;  
 OC Haliotidae; Haliotis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-MUSCLE;  
 RA WATABE S.;  
 RT "Two types of collagen pro alpha chain in the abalone muscle.";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB017600; BAA75668.1; .  
 KW Signal; Collagen.  
 FT SIGNAL 1 18 POTENTIAL.  
 SQ SEQUENCE 1378 AA; 132583 MW; 387BEA97 CRC32;

Query Match 5.6%; Score 134; DB 5; Length 1378;  
 Best Local Similarity 40.4%; Pred. No. 1.89e-04;  
 Matches 19; Conservative 13; Mismatches 14; Indels 1; Gaps 1;  
 Db 801 RSGGIPGNSCTPGDGRAGPPGSGAGGP-PGPGSGATGLSDGGER 846  
 QY 24 RDSGAPGKGVAEPPQKAGPAPGGEAAPPDGGGGGGGLSPGPSDR 70

RESULT 5  
 ID Q00484; PRELIMINARY; PRT; 149 AA.  
 AC Q00484;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
 DE MINI-COLLAGEN PRECURSOR (ISOFORM 1).  
 GN N-COL 1.  
 OS Hydra attenuata (Hydra) (Hydra vulgaris).  
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;  
 OC Hydridae; Hydra.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SF1;  
 RX MEDLINE; 92064646.  
 RA KURZ E.M., HOLSTEIN T.W., PETRI B.M., ENGEL J., DAVID C.N.;  
 RT "Mini-collagens in hydra nematocytes.";  
 RL J. Cell Biol. 115:1159-1169(1991).